# METHODS FOR ESTIMATING PRESENT AND FUTURE INSURGENT STRENGTH

by

Edward Francis Sullivan

United States
Naval Postgraduate School

# THESIS

METHODS FOR ESTIMATING PRESENT AND FUTURE INSURGENT STRENGTH

by

Edward Francis Sullivan

October 1969

# Methods for Estimating Present and Future Insurgent Strength

by

Edward Francis Sullivan
Lieutenant, United States Navy
B.S., Tufts University, 1961

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
October 1969

ABSTRACT

An estimation of the extent of an insurgency appears to be a prerequisite for determining effective counterinsurgency policies. A method is developed to estimate a confidence interval for present insurgent strength based on consumption of a selected commodity. By examining the anticipated effects of dynamic counterinsurgency programs on additions and deletions to insurgent strength, estimates of future insurgent strength can be attempted. For this purpose, recursive relationships are developed describing changes in insurgent strength which occur with changes in level of a single government activity or multiple government activities.

## TABLE OF CONTENTS

| | | | |
|---|---|---|---|
| I. | INTRODUCTION | | 9 |
| II. | MODEL FOR ESTIMATION OF PRESENT INSURGENT STRENGTH | | 12 |
| | A. | NUMERICAL EXAMPLES | 15 |
| | B. | SENSITIVITY ANALYSIS | 20 |
| | C. | EXTENSION OF MODEL TO ESTIMATE OVERALL INSURGENT STRENGTH | 22 |
| III. | MODELS FOR ESTIMATION OF FUTURE INSURGENT STRENGTH | | 24 |
| | A. | INSURGENT RECRUITMENTS | 27 |
| | B. | INSURGENT DEATHS IN COMBAT | 29 |
| | C. | INSURGENT SURRENDERS | 31 |
| | D. | INSURGENT DEATHS FROM DISEASE | 31 |
| | E. | SINGLE GOVERNMENT ACTIVITY MODEL | 34 |
| | F. | MULTIPLE GOVERNMENT PROGRAMS MODEL | 38 |
| IV. | CONCLUSION | | 47 |
| APPENDIX A | | | 51 |
| LIST OF REFERENCES | | | 61 |
| INITIAL DISTRIBUTION LIST | | | 62 |
| FORM DD 1473 | | | 63 |

LIST OF TABLES


| | | Page |
|---|---|---|
| I. | Population and Rice Consumption of Four Villages | 15 |
| II. | Population and Rice Consumption of Ten Villages | 18 |
| III. | Data for Three Government Programs | 45 |
| IV. | Analysis of Variance Table for Deviations from the Mean | 55 |

LIST OF FIGURES

| | | Page |
|---|---|---|
| 1. | Insurgent Recruitments as a Function of Government Activity for a Militarily Active Level of Insurgent Activity ($I_A$) and a Militarily Passive Level of Insurgent Activity ($I_P$). | 28 |
| 2. | Insurgent Deaths in Combat as a Function of Government Activity for a Militarily Active Level of Insurgent Activity ($I_A$) and a Militarily Passive Level of Insurgent Activity ($I_P$). | 30 |
| 3. | Insurgent Surrender as a Function of Government Activity for a Militarily Active Level of Insurgent Activity ($I_A$) and a Militarily Passive Level of Insurgent Activity ($I_P$). | 32 |
| 4. | Insurgent Deaths from Disease as a Function of Government Activity for a Militarily Active Level of Insurgent Activity ($I_A$) and a Militarily Passive Level of Insurgent Activity ($I_P$). | 33 |
| 5. | Insurgent Deaths in Combat as a Function of Government Activity. | 37 |

# I. INTRODUCTION

The presence of insurgents is a problem with which many governments have had to or will have to contend. Since insurgent activity usually represents a political and/or military expression of dissatisfaction with the governmental structure, governments should want to eliminate this activity quickly and completely. An estimation of the extent of the insurgency and in particular the number of insurgents seems to be one of the first steps toward dealing with insurgents. Without some knowledge of insurgent strength, a government appears to be lacking essential information for determining effective counterinsurgency policies.

In the past, insurgent strength has been estimated with varying degrees of accuracy on the basis of government police or troop observations, statements of surrendered or captured insurgents, and insurgent documents which have been discovered or seized. Two specific examples of insurgency were investigated to discover any references to estimated strength. In the Malayan Emergency (1948-1960), government intelligence estimates indicated 12 regiments of insurgents. These were regiments in name only and the strength and organization of the regiments bore no relation to accepted standards.[1] No specific

---

[1] Walker, P. B. G., Notes on the Malayan Emergency: Strategies and Organization of the Opposing Forces, p. 11, Stanford Research Institute, 1967.

estimates of the number of insurgents were given. In the current insurgency in Southern Thailand, insurgent strength is estimated to be between 500 and 700.[2] In the literature concerning both insurgent situations, no reference is made to analytical techniques as a method of estimating insurgent strength.

It is the purpose of this thesis to suggest analytical methods which a government might use to estimate the present and future number of insurgents. In order to proceed logically with this thesis, estimation of present and future insurgent strength will be considered separately.

The method to be suggested for estimating present insurgent strength is based upon consumption of a commodity which is an integral part of the local diet of the area being considered. An examination of the relationship of the consumption of this commodity and the size of the population consuming it should reveal linearity, and knowledge of commodity consumption should provide an estimate of population in the area under consideration. After a commodity consumption variable is decided upon, a confidence interval can be established for the population from a suspected insurgent village. Comparison of the known population with the interval of estimated population, for a particular confidence level, can yield a range of values for the number of insurgents present at the same confidence level. In order for this result to occur, the

---

[2] Hamberg, W. A., and Self, C. R., Communist Terrorist Logistics in Southern Thailand - A Quantitative Analysis, p. 10, Stanford Research Institute, 1968.

known population has to be outside the interval of estimated population. When the known population is in the interval of estimated population, no statement can be made about the presence of insurgents.

There will be two methods suggested for predicting the value of insurgent strength in future time periods. One method of prediction uses past data for the construction of graphs of the various additions and deletions to present insurgent strength, as functions of one government activity. Using these graphs and an estimated level of a particular government activity, future insurgent strength can be predicted. An alternate method for estimating future insurgent strength considers the effects of all government activities on insurgent eliminations and recruitments.

## II. MODEL FOR ESTIMATION OF PRESENT INSURGENT STRENGTH

Since insurgents have to eat to survive and their diet is probably similar to the diet of local non-insurgents, it seemed reasonable to attempt to estimate the present number of insurgents by measuring the amount of a particular food that they consumed. The first determination to be made is the food commodity to be measured. It should be an item which is a necessity for the insurgents either for health purposes or for morale purposes. Depending on the geographical location of the analysis, some possible commodities might be rice, citrus fruit, fish, bread, vegetables and various kinds of meat.

The selected commodity should be one which is consumed by a large proportion of the non-insurgent population and can be purchased by non-insurgents of all income groups. The wholesale sales data for the selected commodity should be used to expedite compilation and to avoid unnecessary possibilities of error. Population data should come from the most recent census.

The estimation of a confidence range of population for non-insurgent villages is predicated upon the assumption that a linear relationship exists between population and commodity consumed. Similarly, suspected insurgent villages are assumed to have linear relationships between population and commodity consumption. If a linear relationship exists between population and commodity consumed, then this relationship can be expressed by an estimated regression equation

$$Y = A + b\,(x - \bar{x}), \tag{1}$$

where Y is the predicted non-insurgent village commodity consumption, x is the known population of a non-insurgent village from the census, and $\bar{x}$ is the mean of the census observations. After differentiating (1) with respect to A and b and setting the results equal to zero, the following relationships exist:

$$A = \frac{\Sigma\, y_i}{k} = \bar{y}, \tag{2}$$

$$b = \frac{\Sigma\,(x_i - \bar{x})\,y_i}{\Sigma\,(x_i - \bar{x})^2} \tag{3}$$

$$\Sigma\,(Y_i - \bar{y})^2 = b^2\;\Sigma\,(x_i - \bar{x})^2, \tag{4}$$

and

$$S^2 = \frac{\Sigma\,(y_i - Y_i)^2}{k - 2} \tag{5}$$

where y is the observed non-insurgent village commodity consumption, k is the number of villages, and $S^2$ is an estimator of $\sigma^2$, the variance of y.[3]

The quantity $\varepsilon$ represents the population of a village estimated on the basis of consumption of a selected commodity. The known popula- of a suspected insurgent village from the census is x'. A confidence interval for the difference between $\varepsilon$ and x' is desired. In the

---

3 A full development of the regression method is presented in Appendix A.

process of presenting expressions for $\varepsilon_1$ and $\varepsilon_2$, which are the endpoints of the confidence interval for $\varepsilon$, a quantity W is defined as

$$W = \frac{t_2^{\,2}\, S^2}{b^2 \, \Sigma\, (x - \bar{x})^2}\,, \tag{6}$$

where $t_2$ is a value from the "student-t" tables. If W has a value less than one-tenth, the expressions for $\varepsilon_1$ and $\varepsilon_2$ are

$$\varepsilon_2 \simeq \bar{x} + \frac{y' - A}{b} - \frac{t_2\, S}{|\,b\,|} \sqrt{(1 + \frac{1}{k}) + \frac{(y' - A)}{b^2 \; \Sigma(x - \bar{x})^2}} \tag{7}$$

and

$$\varepsilon_1 \simeq \bar{x} + \frac{y' - A}{b} - \frac{t_1\, S}{|\,b\,|} \sqrt{(1 + \frac{1}{k}) + \frac{(y' - A)^2}{b^2 \; \Sigma(x_i - \bar{x})^2}}\,, \tag{8}$$

where $y'$ is the observed commodity consumption of a suspected insurgent village.

If $x'$, the known population of a suspected insurgent village from the census, is not in the confidence interval ( $\varepsilon_1$ , $\varepsilon_2$), then ( $\varepsilon_1 - x'$, $\varepsilon_2 - x'$ ) represents the confidence interval for the number of people not counted in the census of the suspected insurgent village. Usually insurgents operate from remote bases located in the jungle or rural areas and it is assumed that insurgents are not counted in the census. If the census was accurate, then ( $\varepsilon_1 - x'$, $\varepsilon_2 - x'$ ) represents a confidence interval for the number of insurgents in the proximity of the suspected insurgent village. If $x'$ is in the confidence interval ( $\varepsilon_1$ , $\varepsilon_2$ ), no statement can be made about the number of insurgents near the suspected insurgent village.

## A. NUMERICAL EXAMPLES

In this section, the applicability of the model is demonstrated through numerical examples using hypothetical data. The assumed scenario is Southeast Asia and the commodity is rice. Rice was selected as the commodity since it exists in the local and insurgent diets. It is assumed that rice is consumed by both insurgents and non-insurgents in approximately equal amounts of one pound per person per day.

For the first example, Table I shows the population and rice consumption figure for four fictitious villages where insurgents are

Table I

Population and Rice Consumption of Four Villages

| Village | $x$ Population | $y$ Rice Consumption (1000 lbs) | $x^2$ | $y^2$ | $xy$ |
|---|---|---|---|---|---|
| Village A | 500 | 178 | 250,000 | 31,684 | 89,000 |
| Village B | 1,200 | 448 | 1,440,000 | 200,704 | 537,600 |
| Village C | 700 | 260 | 490,000 | 67,600 | 182,000 |
| Village D | 1,100 | 397 | 1,210,000 | 157,609 | 436,700 |
| Total | 3,500 | 1,283 | 3,500,000 | 457,597 | 1,245,300 |

assumed not to be present. The village and total values of $x^2$, $y^2$, and $xy$, which are needed to apply the model presented in the beginning of this chapter, are also given in Table I.

Hence, $k = 4$, $\bar{x} = 875$, $\bar{y} = 321$,

$$\Sigma (x_i - \bar{x})^2 = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{k} = 327,500,$$

$$\Sigma (y_i - \bar{y})^2 = \Sigma y_i^2 - \frac{(\Sigma y_i)}{k} = 46,075,$$

and

$$\Sigma (x_i - \bar{x})(y_i - \bar{y}) = \Sigma x_i y_i - \frac{\Sigma x_i \; \Sigma y_i}{k} = 122,675.$$

Using equations (2), (3), (4), and (5), the values of A, b, $\Sigma (Y - \bar{y})^2$, and $S^2$ are

$$A = \frac{\Sigma y_i}{k} = 321,$$

$$b = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = .3746,$$

$$\Sigma (Y_i - \bar{y})^2 = b^2 \; \Sigma (x_i - \bar{x})^2 = 45,952,$$

and

$$S^2 = \frac{\Sigma (y_i - Y_i)^2}{k - 2} = 61.5 .$$

A 95% confidence range for $\varepsilon$ was selected. Therefore,

$$t_{0.975} = -t_{0.025} = 4.303$$

for two degrees of freedom. Using equation (6),

$$W = \frac{t_2 \; S^2}{b^2 \; \Sigma (x - \bar{x})^2} = .024 \quad .$$

Suppose the population of a suspected insurgent village (x') was 1300 (from census data) and the annual rice consumption (y') was 510,000 pounds. Since $W < 0.1$, approximations (7) and (8) can be used and they yield $\varepsilon_2 = 1508$ and $\varepsilon_1 = 1252$. Since $x' = 1300$ is inside the 95% confidence interval for $\varepsilon$, (1252, 1508), no statement can be made about the presence of insurgents in the suspected insurgent village. Similarly, for $y' = 510{,}000$, 90% confidence limits on $\varepsilon$ yield the interval (1293, 1467) and no statement can be made about the presence of insurgents in the suspected insurgent village.

In the second example, the number of fictitious villages where insurgents are assumed not be present is increased to ten. It is hoped that the additional data from these villages may provide a smaller 95% confidence interval for $\varepsilon$ and lead to a conclusive statement concerning the number of insurgents in the suspected insurgent village.

Table II shows the population and rice consumption figures for ten fictitious villages where insurgents are assumed not to be present. The village and total values of $x^2$, $y^2$, and xy, which are needed to apply the model presented in the beginning of this chapter, are also given in Table II.

Hence $k = 10$, $\bar{x} = 1{,}020$, $\bar{y} = 373$,

$$\Sigma (x_i - \bar{x})^2 = 1{,}667{,}800,$$

$$\Sigma (y_i - \bar{y})^2 = 222{,}628,$$

and

$$\Sigma (x_i - \bar{x})(y_i - \bar{y}) = 609{,}020.$$

Using equations, (2), (3), (4), and (5), the values of A, b, $\Sigma (y_i - \bar{y})^2$, and $S^2$ are

Table II

Population and Rice Consumption of Ten Villages

| Village | x<br>Population | y<br>Rice Consumption (1000 lbs) | $x^2$ | $y^2$ | xy |
|---|---|---|---|---|---|
| Village A | 500 | 178 | 250,000 | 31,684 | 89,000 |
| Village B | 1,200 | 448 | 1,440,000 | 200,704 | 537,600 |
| Village C | 700 | 260 | 490,000 | 67,600 | 182,000 |
| Village D | 1,100 | 397 | 1,210,000 | 157,609 | 436,700 |
| Village E | 350 | 125 | 122,500 | 15,625 | 43,750 |
| Village F | 880 | 326 | 774,400 | 106,276 | 286,880 |
| Village G | 1,320 | 480 | 1,742,400 | 230,400 | 633,600 |
| Village H | 950 | 349 | 902,500 | 121,801 | 331,550 |
| Village I | 1,500 | 550 | 2,250,000 | 302,500 | 825,000 |
| Village J | 1,700 | 615 | 2,890,000 | 378,225 | 1,045,500 |
| Total | 10,200 | 3,728 | 12,071,800 | 1,612,424 | 4,411,580 |

$$A = \frac{\Sigma Y_i}{k} = 373,$$

$$b = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = .3652,$$

$$\Sigma = (Y_i - \bar{y})^2 = b^2 \, \Sigma (x_i - \bar{x})^2 = 222,392,$$

and

$$S^2 = \frac{\Sigma (y_i - Y_i)^2}{k - 2} = 29.5 \quad .$$

A 95% confidence range for $\varepsilon$ was selected. Therefore,

$$t_{0.975} = -t_{0.025} = 2.306$$

for eight degrees of freedom. Using equation (6),

$$W = \frac{t_2 S^2}{b^2 \Sigma (x_i - \bar{x})^2} = .0007 \quad .$$

The suspected insurgent village to be used in this example is the same as the one used in the first example. There, x' = 1300 and y' = 510,000 pounds. Since $W < 0.1$, approximations (7) and (8) can be used and they yield $\varepsilon_2 \simeq 1440$ and $\varepsilon_1 \simeq 1366$. Since x' = 1300 is outside the 95% confidence interval for $\varepsilon$, (1366, 1440), then (66,140) represents a 95% confidence interval for the number of insurgents near the suspected insurgent village. Similarly, 90% confidence limits on $\varepsilon$ yield the interval (1373, 1433) and the 95% confidence interval for the number of insurgents near the suspected insurgent village is (73,133).

The foregoing examples were used to illustrate the application of the model for estimating present insurgent strength. After a commodity consumption variable is decided upon, in accordance with the necessary assumptions, a confidence interval can be established for the population from a suspected insurgent village. Comparison of the known population with the interval of estimated population, for a particular confidence level, can yield a range of values for the number of insurgents present at the same confidence level. In order for this result to occur, the known population has to be outside the interval of estimated population.

In the first example, the supposed population, 1300, was within both the 95% and 90% confidence intervals of the estimated population. In the second example, 1300 people was less than the lower limit of the interval of estimated population at both the 95% and 90% confidence levels and definite statements about the presence of insurgents could be made.

These examples also indicate the effect of changing the number of villages considered. The population and rice consumption of Village A through Village D was the same in both examples. Adding six more villages in the second example resulted in definite statements about the presence of insurgents for both confidence levels.

## B. SENSITIVITY ANALYSIS

Generally, one would want the confidence interval for $\varepsilon$ to be as small as possible. The expressions for $\varepsilon_1$ and $\varepsilon_2$ were

$$\varepsilon_2 \simeq \bar{x} + \frac{y' - A}{b} - \frac{t_2 S}{|b|} \sqrt{\left(1 + \frac{1}{k}\right) + \frac{(y' - A)^2}{b^2 \Sigma (x_i - \bar{x})^2}} \tag{7}$$

and

$$\varepsilon_1 \simeq \bar{x} + \frac{y' - A}{b} - \frac{t_1 S}{|b|} \sqrt{\left(1 + \frac{1}{k}\right) + \frac{(y' - A)^2}{b^2 \Sigma (x_i - \bar{x})^2}} \tag{8}$$

An examination of the approximations (7) and (8) reveals that these approximations are sensitive to changes in the difference between the observed commodity consumption of a suspected insurgent village and the mean commodity consumption of non-insurgent villages $(y' - A)$, the number of non-insurgent villages, and of course, the confidence level. The effects of changes in $(y' - A)$ and $k$ will now be discussed.

The size of selected non-insurgent villages can effect the size of the confidence interval for $\varepsilon$. By selecting non-insurgent villages in such a way that their mean commodity consumption (A) is relatively close to $y'$, the value of the expression $(y' - A)^2$ will be small. As the value of $(y' - A)^2$ becomes smaller, the confidence interval of $\varepsilon$ becomes smaller.

The number of villages to be considered (k) has a three-way effect on what the length of the confidence interval for $\varepsilon$ will be. More observations will result in the expression $(1 + \frac{1}{k})$ being smaller, the value of $t$ will decrease as the degrees of freedom $(k - 2)$ becomes larger, and the standard deviation (s) should decrease as the sample size becomes larger. The length of the confidence interval changes inversely with changes in $k$, as noted in the previous section where $k$ changed from four to ten.

In the sensitivity analysis above, it has been shown that the size and number of non-insurgent villages selected and the confidence level desired will have an effect on the size of the confidence interval of $\varepsilon$. Prior to performing computations to estimate the number of insurgents in a village, the results of the sensitivity analysis should be considered to assist in controlling the range of $\varepsilon$.

## C. EXTENSION OF MODEL TO ESTIMATE OVERALL INSURGENT STRENGTH

The examples in Section A were confined to estimations of insurgent strength in individual villages. This limited analysis may be useful to a government desiring information about insurgent strength for a particular village. Usually a government is more interested in the overall strength of an insurgent force.

There are two methods which can be used to extend the model to estimate overall insurgent strength. The confidence interval obtained from these two methods depends on the number and size of the non-insurgent villages which will be observed.

The first method would be to observe the population and commodity consumption of a suspected insurgent area. If the area in which insurgents were thought to be present has a greater population than any of the non-insurgent villages used to estimate the regression line, then the value of commodity consumption for the entire insurgent area ($y'$) would be large in relation to the mean commodity consumption of non-insurgent villages ($A$). From the sensitivity analysis in Section B, it is known that a large value for the expression $(y' - A)^2$ can result in a large confidence interval for $\varepsilon$. Since a large number of non-insurgent villages ($k$) could be observed using this method, the effect of a large $(y' - A)^2$ on the size of the confidence interval for $\varepsilon$ will be counteracted to some degree.

A second method also would consider the population and commodity consumption of a suspected insurgent area. Non-insurgent villages would be grouped into areas approximately the same size as the suspected insurgent area, if this is possible. The value of $y'$ and A would then be approximately equal so that the value of the expression $(y' - A)^2$ would be small. By grouping the observations from non-insurgent villages, k would be reduced so the effect of a small $(y' - A)^2$ on the size of the confidence interval for $\varepsilon$ is counteracted by a small value of k.

If enough time and resources are available, both these methods should be used and the smallest confidence interval for $\varepsilon$ should be chosen. If time and resources are limited, the second method involves fewer computations and should be used.

## III. MODELS FOR ESTIMATION OF FUTURE INSURGENT STRENGTH

In the preceding chapter, a model was developed to find confidence intervals for the present number of insurgents in a country. It is desired also to have procedures to estimate the number of insurgents at some future time. In the following paragraphs, some of the more important factors causing change in insurgent strength will be considered and their individual influences on overall insurgent strength investigated.

Additions to the number of insurgents can occur from child births (within the insurgent community) and recruitments. An eighteen year time lag probably exists before a child can be classified as a useful insurgent. When a non-insurgent voluntarily becomes an insurgent, a recruitment occurs. Insurgent propaganda and/or military operations are usually the methods employed to solicit recruitments. Insurgents are considered to be engaged in insurgency on a full-time basis, so insurgent sympathizers and those people forced to provide help for the insurgents through coercion are not considered insurgents.

Depletions to insurgent strength can occur from deaths in combat, surrenders, deaths from disease, and retirements (old age). Deaths in combat include insurgents who are captured and it is assumed that these personnel will be immobilized for the duration of the conflict. Surrenders include deserters. Insurgent deaths from disease usually result from malnutrition, starvation, or infection. When an insurgent is wounded in combat, escapes capture and later dies from his wound, his death will be classified as a death from disease.

Each of the additions or depletions to insurgent strength, as defined above, is a function of many variables. Two of the more important of these variables appear to be government activity and insurgent activity. Analysis of insurgent recruitments, deaths in combat, surrenders, and deaths from disease, as functions of government activity and insurgent activity, will be attempted in this chapter.

Insurgent births and retirements, although functions of government and insurgent activity, are probably affected, to a a greater degree, by the age distribution of the insurgent population and the length of the conflict. More births would be expected in a group of youthful insurgents and more retirements would be expected in a group of older insurgents. It would seem reasonable that the longer the conflict the more consideration that should be afforded to births and retirements. For the present analysis, however, it will be assumed that the collective impact of births and retirements on future insurgent strength is negligible.

Generally, definitions of government activity in combating insurgency can fall into either of two categories; military or non-military. Examples of definitions of military government activity levels could be "government troop-days spent on patrol per month" and "the number of tons of bombs dropped by the government per month." Examples of definitions of non-military government activity levels could be "the number of pounds of government literature disseminated to suspected insurgent-occupied villages per month," "the number of inhabitants, in an area where the presence of insurgents is suspected, treated by mobile government medical units per month," and "the monthly value, in local currency, of village and commodity support furnished by government-sponsored activities in insurgent areas." Although any of these definitions of government activity could be appropriate in analyzing the distributions of additions and deletions to insurgent strength, "government troop-days spent on patrol per month" has been chosen as the index of government military activity to be used in this analysis.

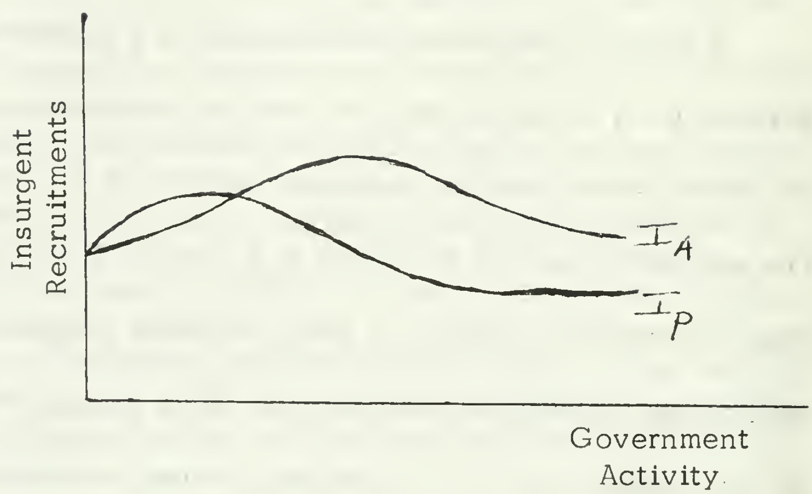
Since an index of government military activity has been chosen, it seems reasonable to also review insurgent activity in terms of its military nature. The approach taken will be essentially speculative and considerations of insurgent activity will be limited to the ends of the spectrum of military activity: militarily passive and militarily active. Communists in South Thailand have in recent years behaved as militarily passive insurgents and Castro rebels in Cuba serve as a past example of the militarily active insurgents.

Data, consisting of the number of government troop-days spent on patrol for monthly periods and the number of insurgent births, recruitments, deaths in military engagements, surrenders, deaths from disease, and retirements for quarterly periods, would be needed to determine the relationships among additions and deletions to insurgent strength and different levels of government activity. At this present time, this data does not appear to exist for the Thailand example. In the absence of actual data, examples of possible functional relationships between additions and deletions to insurgent strength and government and insurgent activity will be hypothesized and discussed. These examples form the rationale for later development of a recursive model for estimating future insurgent strength.

## A. INSURGENT RECRUITMENTS

A graph of insurgent recruitments as a function of government activity for a militarily passive level of insurgent activity ($I_P$) and a militarily active level of insurgent activity ($I_A$) is shown in Figure 1. The militarily passive insurgent ($I_P$) curve will be considered first. When government activity is low, insurgent propaganda concerning government oppression probably has little impact on the populace but as government activity is increased, there is evidence to support claims of governmental oppression and recruitments increase. When government activity is increased to the point that the insurgent propaganda activities are seriously impaired, insurgent recruitments should cease to increase and remain constant. If government activity is further increased, the appeal of the insurgent cause should decrease as a result of increased danger and recruitments probably will decrease.

As government activity is increased, it seems reasonable that the engagements between government troops and militarily active insurgents should increase. Since the insurgents should be successful in combat at lower levels of government activity, recruitments should




Insurgent Recruitments as a Function of Government Activity for a Militarily Active Level of Insurgent Activity ($I_A$) and a Militarily Passive Level of Insurgent Activity ($I_P$)

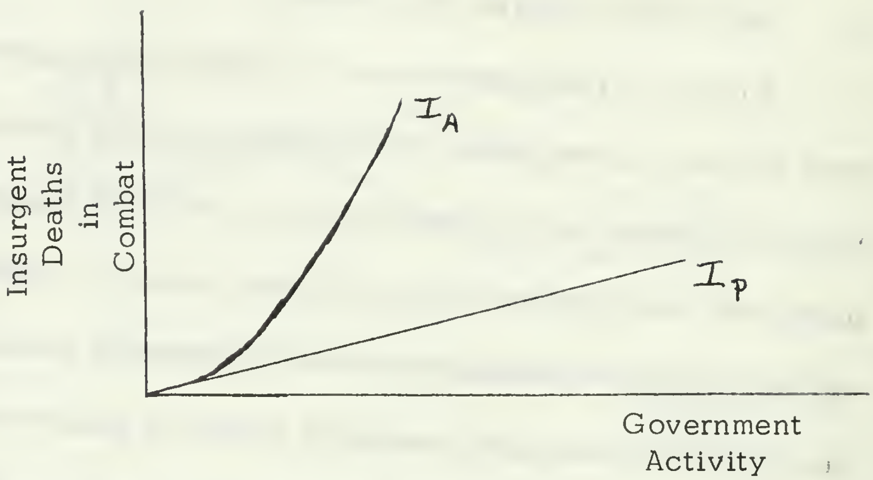
Figure 1

increase. A further increase in government activity could check the insurgents success and recruitments would remain constant since people are uncertain of the outcome of the conflict. Additional government activity probably will relegate the insurgent to hit and run tactics since he has to exercise more caution before engaging government troops in combat. Interest and support of the insurgent should subside at this level of government activity and recruitments should decrease.

## B. INSURGENT DEATHS IN COMBAT

A graph of insurgent deaths in combat as a function of government activity for two levels of insurgents activity is shown in Figure 2. When insurgents are militarily passive, they will probably attempt to avoid any confrontation with government troops. Deaths in combat for militarily passive insurgents are those insurgent deaths which occur in combat confrontations instigated solely by government troops. Since this level of insurgent activity revolves around propaganda, increased government activity probably will force the insurgents to become clandestine in their activities but the additional government patrolling should result in more insurgent discoveries, captures, and deaths.

As government troop-days patrolling per month are increased, engagements with militarily active insurgents should increase. As engagements increase, insurgents probably will suffer increased deaths in combat. The $I_A$ curve has a steeper slope than the $I_P$ curve since open combat exposes more insurgents to death and capture.




Insurgent Deaths in Combat as a Function of Government Activity for a Militarily Active Level of Insurgent Activity ($I_A$) and a Militarily Passive Level of Insurgent Activity ($I_P$)

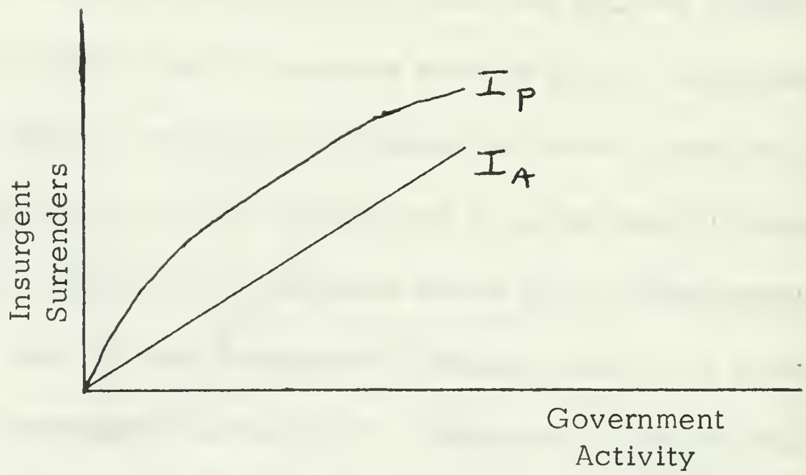
Figure 2

## C. INSURGENT SURRENDERS

A graph of insurgents surrenders as a function of government activity for the two extremes of insurgent activity is shown in Figure 3. Militarily passive insurgents will concentrate on propaganda and probably will attempt to avoid combat with the government. Usually, they will not be as well-disciplined and dedicated to their cause as militarily active insurgents, who probably are more concerned with military training and who risk their lives in battle. Militarily active insurgents should be more successful than militarily passive insurgents in securing food at all levels of government activity since they are probably more prone to use force in order to satisfy food needs. As a consequence of the above statements, government activity should result in a larger number of surrenders from $I_P$ than $I_A$. However, for both levels of insurgency, increases in government activity could impose physical hardships through problems in logistics, exposure to elements, and increased vigilance (loss of sleep). As these hardships and suffering increase, insurgent surrenders should increase.

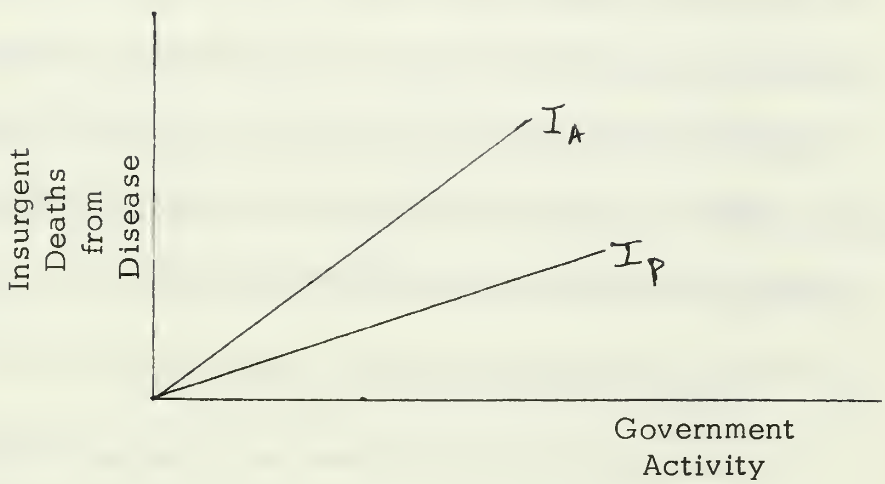
## D. INSURGENT DEATHS FROM DISEASE

A graph of insurgent deaths from disease as a function of government activity for two levels of insurgent activity is shown in Figure 4. Assuming insurgents are operating from a jungle or rural environment, increases in government activity should cause insurgent exposure to cuts, poor diet, starvation, fatigue, and wounds to increase. Insurgent access to food and medical supplies should become more difficult.




Insurgents Surrenders as a Function of Government Activity for a Militarily Active Level of Insurgents Activity ($I_A$) and a Militarily Passive Level of Insurgent Activity ($I_P$)

Figure 3




Insurgent Deaths from Disease as a Function of Government Activity for a Militarily Active Level of Insurgent Activity ($I_A$) and a Militarily Passive Level of Insurgent Activity ($I_P$)

Figure 4

Therefore, increases in government activity should result in increases in insurgent deaths from disease.

Militarily active insurgents will probably have a greater incidence of combat wounds and thus should suffer more deaths from disease than militarily passive insurgents. Accordingly, the $I_A$ curve should be above the $I_P$ curve.

Since there is normally a token amount of government police personnel assigned to areas of insurgency, a zero level of government activity in the preceding graphs would be very unlikely. This should be considered when looking at the graphs in Figure 1 through Figure 4. For instance, it is possible that there could be some insurgent surrenders when there is no government activity.

## E. SINGLE GOVERNMENT ACTIVITY MODEL

Actual data for the additions and deletions to insurgent strength and levels of government activity may be available for current emergencies but there does not appear to be any evidence that this information is being used to estimate future insurgent strength. After appropriate reporting procedures are established for government military agencies, records of government activity, measured in patrol troop-days per month, insurgent surrenders per quarter, and insurgent deaths in combat per quarter could be established and maintained. Less precise records of insurgent recruitments and deaths from disease could be established and maintained from information received from government intelligence agencies.

Once actual data is available, graphs can be plotted and analyzed. The method suggested for construction of these graphs requires estimated quarterly totals of insurgent recruitments, deaths in combat, surrenders, and deaths from disease for the actual average monthly government patrol troop-days during the same quarter. As an example, assume actual insurgent surrenders during a particular quarter were 100 and the actual monthly government patrol troop-days were 600, 840, and 720. Then a point (X,Y) on the graph of insurgent surrenders would have the value (730, 100). If data for two years is used, then each of the graphs of quarterly insurgent recruitments, deaths in combat, surrenders, and deaths from disease, plotted against average monthly government patrol troop-days, will have eight data points. If continuity is assumed, a curve can be drawn through the eight data points.

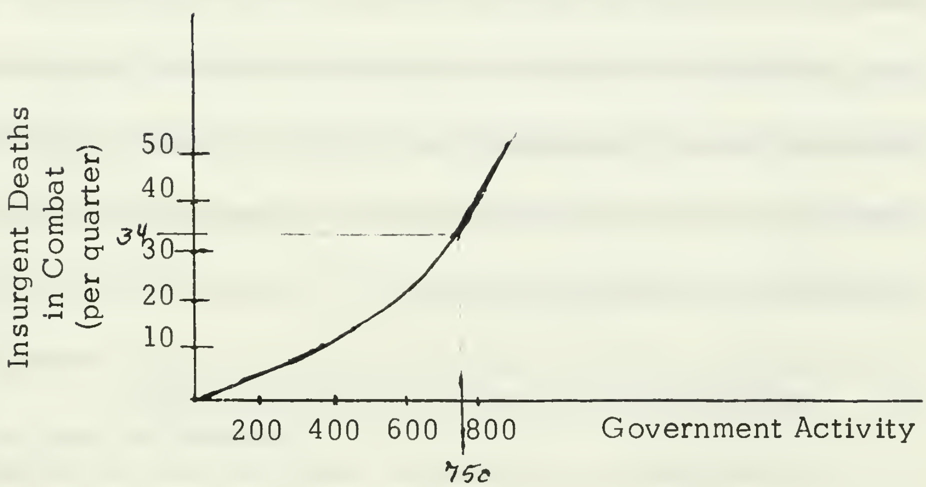
The model to estimate future insurgent strength is based on these graphs. In order to have a dynamic model to accommodate and reflect changes in insurgent and government tactics, the curves, described above, would have to be updated by incorporating the most recent quarterly observation and deleting the oldest quarterly observation. If the curves are constructed in this manner, they will be comprised of the most recent two years of observations.

Since government activity is a controlled variable and therefore can be predicted approximately, the predicted average monthly government patrol troop-days can be used to estimate quarterly additions and deletions to insurgent strength. By examining the graphs constructed from past data, the intersection of the curve of insurgent deaths from combat, for example, and the predicted average monthly level of government activity will yield a predicted value for quarterly insurgent deaths in combat. As an example, assume that the curve in Figure 5 has been constructed by using past quarterly figures for insurgent deaths in combat and average monthly government patrol troop-days. If the average monthly level of government activity was predicted to be 750 patrol troop-days for the next quarter, insurgent deaths in combat would be estimated at 34 for the next quarter.

If the curves of insurgent recruitments, surrenders, and deaths from disease are examined in the same manner for a predicted average monthly level of government activity of 750 patrol troop-days, the aggregate predicted addition or deletion to insurgent strength for next quarter can be estimated.

In a protracted insurgency, insurgent births and retirements could be included in the analysis. In a given situation, some information probably would be available concerning the insurgent age distribution and length of the conflict. If available, this information could be used to estimate insurgent births and retirements.

When government resources are limited, a particular range of future insurgent strength can be selected and the model can be used in reverse to see what level of government activity would be needed to contain the future numbers of insurgents within this range. At this




Insurgent Deaths in Combat as a Function of Government Activity

Figure 5

point, it should be reiterated that insurgent growth or decline is a function of many things, of which government activity is assumed predominant.

There are obvious arguments against the accuracy of estimates provided by this model. Some of these arguments could be that the assumption of continuity may not be valid for the areas of the curves between the data points provided by past observations, that the extreme ends of the curves are the product of questionable extrapolation, and that the curves are not responsive enough to current changes in tactics by either the government or the insurgents. All of these arguments are valid but the information provided by this model should be better than no information at all.

## F. MULTIPLE GOVERNMENT PROGRAMS MODEL

In the previous discussion, only one index of government activity was used to estimate future insurgents strength. As was stated previously, various military and non-military government programs are usually undertaken simultaneously in a counterinsurgency effort. In this section, a model will be developed to relate the impact of multiple government programs on future insurgent strength. First, a single government program will be considered in the development of a model and then the model will be expanded to include multiple government programs.

The population is assumed to be divided into distinct groups of insurgents and non-insurgents. Let $p_k$ be the fraction of the total population which is insurgent at the beginning of time period k; then the term $(1 - p_k)$ will be the fraction of the total population which is non-insurgent at the beginning of time period k.

Let $G_k$ be the fraction of total population "reached" by any government program in time period k. The word "reached" is used to denote exposure to any government program. An individual may be "reached" by receiving a government leaflet or by receiving medical attention which saves his life, so there is no distinction made between the degree of government effort required to "reach" an individual. The scope is the only characteristics of a government program represented by $G_k$. Let $\eta$ be the conditional probability that an insurgent is "eliminated" if he is "reached." The word "eliminated" is used to denote the removal of an insurgent from the insurgent population by surrender or death or capture.

Let $I_k$ be the fraction of non-insurgents "reached" by insurgents in time period k. $I_k$ is determined independently of any government program. Let $\rho$ be the conditional probability that a non-insurgent is "converted" to insurgency if he is "reached." The word "converted" is used to denote the migration of a non-insurgent from the non-insurgent population to the insurgent population.

If it is assumed that the insurgents are randomly scattered throughout the total population, the expression $p_k G_k$ is the fraction of population which is insurgent and "reached" by the government in time k. The term $p_k G_k \eta$ is the fraction of population which is insurgent, "reached" by the government, and "eliminated" by the government during time period $k$. The term $(1 - p_k)\ I_k$ is the fraction of population which is non-insurgent and "reached" by the insurgents in time period $k$. The expression $(1 - p_k)\ I_k\ \rho$ is the fraction of population which is non-insurgent, "reached" by the insurgents, and "converted" to insurgency during time period $k$.

Let $p_{k+1}$ be the fraction of the total population which is insurgent at the beginning of time period $(k + 1)$. Then $p_{k+1}$ will be equal to $p_k$ minus the fraction of insurgents "eliminated" plus the fraction of non-insurgents "converted" during time period $k$. This relationship can be expressed as follows:

$$p_{k+1} = p_k - p_k G_k\ \eta + (1 - p_k)\ I_k\ \rho\ . \qquad (9)$$

In (9), $\eta$ reflects government efficiency in "eliminating" insurgents and $\rho$ reflects insurgent efficiency in "converting" non-insurgents. Since government activity not only results in fractions of the population being "reached" by the government but has some effect on fractions of the population "reached" by the insurgents, $\eta$ and $\rho$ might be considered as functions of $G_k$. Assuming that $\eta$ and   are linear functions of $G_k$, they could have the forms $(\eta_o + \beta_2\ G_k)$ and $(\rho_o + \beta_2\ G_k)$ respectively. In this context $\eta_o$ would represent the probability that an insurgent is eliminated when no insurgents are reached by government. The term $\rho_o$ represents the fraction of non-insurgents converted to insurgency in period $k$ if reached by the insurgents when no government programs exist. The terms $\beta_1$ and $\beta_2$ represent the program efficiency slopes.

If an extreme government program requires armed pursuit and ultimate capture of insurgents, it should be successful in the sense that insurgents will be eliminated but the effects of such a program on non-insurgents may contribute to their successful recruitments by insurgents. Conversely, a government good-will program which is directed toward the entire population in the hope of winning over insurgents peacefully may not be very successful in eliminating militant insurgents but may serve to greatly hamper insurgent recruitments.

It is assumed that if government efficiency in removing insurgents ( $\eta$) increases with the slope $(G_k)$ of a program, then insurgents recruiting efficiency ( $\rho$) will be reduced equally. Therefore, in this analysis, the slopes of the lines ( $\eta_o + \beta_1 G_k$) and ( $\rho_o + \beta_2 G_k$) will be equal in value but will have opposite signs, i.e., $\beta_1$ will be equal to $-\beta_2$ . Now (9) can be restated as

$$p_{k+1} = p_k - p_k G_k (\eta_o + \beta G_k) + (1 - p_k) I_k (\rho_o - \beta G_k). \quad (10)$$

It is noted that (10) is a quadratic equation in $G_k$ and represents a recursive relationship for the fraction insurgent, of the population. The term $p_k G_k (\eta_o + \beta G_k)$ represents the fraction of insurgents "eliminated" by the government during time period k. The term $(1 - p_k) I_k (\rho_o - \beta G_k)$ is a cross product of $I_k$ and $G_k$ and represents the fraction of non-insurgents "converted" to insurgents. This term reflects the effects of government programs on insurgent recruiting efficiency.

A government will usually have many programs that have insurgent "elimination" as their ultimate goal. Some examples of military and non-military government activity were given at the beginning of this chapter. Any or all of these government programs can be utilized in a particular time period and it would be desirable to modify the recursion of (9) to reflect the effects of simultaneous but different government programs.

Let $G_k^i$ be the fraction of total population "reached" by the $i^{th}$ government program during time period $k$ and $\eta^i$ be the conditional probability that an insurgent is eliminated if he is "reached" by the $i^{th}$ government program. The term $p_k G_k^i \eta^i$ is the fraction of population which is insurgent, "reached" by government program $i$, and "eliminated" by government program $i$ during the time period $k$. If all programs were mutually exclusive in their scope, equation (9) could be modified to include all government programs by summing the second term of (9) over all $i$ programs. However, it seems reasonable that some insurgents may be "reached" by more than one government program. For example, if an insurgent was exposed to two programs, the probability of his elimination would be greater than it would have been if he was "reached" by just one of the programs. These interactions will be considered below.

The term $(1 - p_k G_k^i \eta^i)$ is the fraction of population which is not (1) insurgent, (2) "reached" by government program $i$, and (3) "eliminated" by government program $i$ during time period $k$. The product $\pi\,(1 - p_k G_k^i \quad \eta^i)$ is the fraction of population which is not (1) insurgent, (2) "reached" by any government programs, and (3) "eliminated" by <u>any</u> programs in time period k. Therefore, the term $[1 - \pi\,(1 - p_k G_k^i \quad \eta^i)\,]$ is the fraction of population which is insurgent, "reached," and "eliminated" for all of the government programs during time period k. After substituting $\eta_o + \beta^i G_k^i$ for $\eta^i$ in the last expression, it becomes

$$[\,1 - \pi \;\{\,1 - p_k G_k^i \;(\,\eta_o + \beta^i G_k^i)\;\}\;]\,.$$

This expression replaces the second term on the right hand side of (10) when multiple government programs exist.

The third term on the right hand side of (10),

$$[\,(1 - p_k)\; I_k\;(\rho_O - \beta\, G_k)\,]$$

measured the effect of a government program on recruitment of non-insurgents by insurgents. The term $\rho_O$ was defined as the probability that a non-insurgent is "converted" to insurgency if he is "reached" by the insurgents when there are no government programs. The effects of any government program on insurgent recruitments would probably not be instantaneous and will be assumed to lag one time period. Therefore, insurgent recruiting efficiency will be affected by government programs in time period (k - 1). The effect of any government program on insurgent recruitments was assumed to be $-\beta G_k$ and the effect of all government programs on insurgent recruitments is

$$\bar{\beta}\ [1 - \pi\ (1 - G^{i}_{k-1})]\ ,$$

where

$$\bar{\beta} = \frac{\Sigma\ \beta^{i} G^{i}_{k-1}}{\Sigma\ G^{i}_{k-1}}$$

and the term in brackets is the fraction of the population reached by all government programs. The expression

$$[(1 - p_k)\ I_k \quad \{\rho_o - \bar{\beta}\ [1 - \pi\ (1 - G^{i}_{k-1})]\ \}]$$

will replace the third term on the right hand side of (10) when multiple government programs exist.

Replacing the last two terms of (10) by the expressions which were developed to recognize multiple government programs, the expression for the fraction of the total population that is insurgent at the end of time period k becomes

$$p_{k+1} = p_k - [\ 1 - \pi\ \{1 - p_k G^{i}_{k}\ (\eta_o + \beta^{i} G^{i}_{k})\ \}] \qquad (11)$$
$$+ (1 - p_k)\ I_k\ [\rho_o - \bar{\beta} \cdot \{\ 1 - \pi\ (1 - G^{i}_{k-1})\ \}\ ]\ .$$

As an example of the computation of $p_{k+1}$ will be given, it is assumed that three government programs are in existence during time periods (k - 1) and k. The first program is a military one which requires searching of homes and villages for insurgents. The second program is non-military and requires government propaganda leaflets to be distributed to villages by air. The third program is also non-military and involves sending medical assistance teams to villages to treat disease.

Hypothetical data for the three programs is presented in Table III. The fraction of the population "reached" by the first program is observed to be about 1/25 in time period (k - 1) and is estimated to increase to 1/20 in time period k. The fraction of the population "reached" by the second program is observed to be about 1/4 in time period k. The fraction of the population "reached" by the third program is observed to be about 1/8 in time period (k - 1) and is estimated

Table III

Data for Three Government Programs

| | Program One | Program Two | Program Three |
|---|---|---|---|
| Scope of Program i during time period k | 1/20 | 1/4 | 1/10 |
| Scope of Program i during time period k-1 | 1/25 | 1/4 | 1/8 |
| Program efficiency slope during any time period | -3/4 | 1/10 | 1/4 |

to decrease to 1/10 in time period k. Program efficiencies slopes for any time period are estimated to be -3/4, 1/10, and 1/4 for the first, second, and third programs respectively. The fraction of population which is insurgent at the beginning of time period k ($p_k$) is estimated to be 1/10. The fraction of non-insurgents "reached" by the insurgents during time period k ($I_k$) is estimated to be 1/5. When no government programs exist, the elimination efficiency ($\eta_o$) and recruiting efficiency ($\rho_o$) are both estimated to be 1/8. When these values are substituted into equation (11),

$$p_{k+1} \sim \frac{1}{10} - \frac{3}{500} + \frac{4}{500} \sim \frac{51}{500} \quad .$$

When using equation (11), $p_k$, the fraction of the total population which is presently insurgent could be estimated by the model in Chapter II. The $G_k^i$ expressions should not be difficult to estimate since the fraction of the population reached by a program may, in most cases, be reasonably well-estimated. The estimation of the $i^{th}$ program efficiency slope ($\beta^i$), the elimination efficiency when no government programs exist ($\eta_o$), the recruiting efficiency when no government programs exist ($\rho_o$), and the fraction of non-insurgents "reached" by the insurgents during time period k ($I_k$) will be more difficult but should be feasible.

The single government activity model was very sensitive to the validity of the assumption that continuous curves of additions and deletions to insurgent strength could be drawn for various levels of a single index of government activity. The multiple government programs model considers the effect of all government programs on insurgent elimination and recruitments. It offers an alternative when the use of the single government activity model might be inappropriate. After examining a particular insurgent scenario, the more appropriate model for estimating insurgent strength should be chosen.

## IV. CONCLUSION

In the preceding chapters models for estimating present and future strength have been presented. The model for estimating present insurgent strength should be an improvement on guesswork. It should be applicable in most areas of the world if the commodity selected is a staple item in the local diet. It is possible that commodity consumption data for insurgent or non-insurgent villages may be affected by the age and sex mix of the corresponding populations. By weighting age and sex to modify population data, it should be possible to preserve linearity between population and commodity consumption.

Future application of the model for estimating present insurgent strength in an actual insurgency situation would be desirable, since the utility of the model could be tested. If actual figures for populations and applicable commodity consumption are available and the necessary assumptions are fulfilled, a confidence interval for present insurgent strength could be found.

Since force ratios are usually a consideration in planning the number of government troops to meet an insurgent threat, knowledge of estimated insurgent strength would assist in determining this number. As an example, if a ten-to-one force ratio of government troops to insurgent is desired and a 95% confidence interval for insurgent strength is (875, 935), the number of government troops assigned to cope with insurgents should be in the interval (8750, 9350).

The first model for estimating future insurgent strength, the single government activity model, requires historical data concerning insurgent additions and deletions for levels of a particular government activity. It has been suggested that two years of data would be sufficient. It is a dynamic model in the sense that the most current data replaces the oldest data. If a range of values for average monthly government activity exists, continuous curves of insurgent additions and eliminations as functions of a government activity can be drawn.

If the continuity assumption could be satisfied and minimum extrapolation performed, this model should provide reasonable predictions. An area of future work could be using the model for estimating future insurgent strength and evaluating its predictions. This could be done using data which is available from a past insurgency or amassing relevant data from a current insurgency.

The second model for estimating future insurgent strength, the multiple government program model, uses $G_k$, the fraction of the population reached by the government in time period $k$, as a decision variable. The effects of all programs rather than a single program are considered in order to estimate insurgent eliminations and recruitments. This model requires the observations of the fraction of the population reached by government program $i$ ($G^i$) from the preceding period and estimates of $G^i$ for the current period. In this respect, the multiple government program model can be useful alternative for the single government activity model when two years of historical data are not available.

It was noted that the multiple government programs model was recursive. Therefore, by n repeated applications of the recursive relationship, the fraction of population which is insurgent at the beginning of time period $k+n$ $(p_{k+n})$ could be estimated. In this context, this model could be used as a planning tool for estimating the effectiveness of an overall government programs effort.

Future work could be done in some of the areas covered in this thesis. In the model for estimating present insurgent strength, a linear relationship was assumed to exist between population and commodity consumption. It was suggested that some type of food be selected as the commodity to be measured. Future investigation could uncover a better commodity than food to use in this type of model. Possible substitutes might be medical supplies or gasoline.

In this single government activity model, the selection of a good index of government activity is necessary. This index should be one for which functional relationships exist with additions and deletions to insurgent strength. Several possibilities may have to be considered before discovering a satisfactory index.

In the multiple government program model, the slopes of the lines $(\eta_o + \beta, G_k)$ and $(\rho_o + \beta_2 G_k)$ were assumed to be equal in value but had opposite signs. When government efficiency in removing insurgents $(\eta)$ is increasing with the scope $(G_k)$ of a program, then the effects of this program on insurgent recruiting efficiency $(\rho)$ should be examined. Insurgent recruiting efficiency $(\rho)$ should be reduced by the same amount that $n$ is increased. By this type of investigation, the validity of the assumption $\beta_1 = -\beta_2$ could be established.

It is hoped that the discussion, observations, and modeling techniques contained in this thesis will benefit future counterinsurgency analysis. Although the problem of quantifying and dealing with the many variables inherent to counterinsurgency analysis has by no means been completely solved, it is felt that a start has been made.

APPENDIX A

In this appendix, a model to estimate linear regression is developed and extended to find intervals for $\varepsilon$ , the true value of x, for any confidence level. In this thesis, x represents the known population of a village from the census and $\varepsilon$ represents the population of villages based on commodity consumption. The objective of the method is to find a confidence interval for the difference between $\varepsilon$ and x.

Assume y is distributed normally about an expected value $\theta$ with variance $\sigma^2$ and all observations are independent. The number of observations is defined as k. Assume that

$$\theta = \alpha + \beta (x - \bar{x}) , \tag{12}$$

a simple linear function of x.[4] Sample estimates A, b, and $S^2$ of $\alpha$ , $\beta$ , and $\sigma^2$ , are desired as well as the distributions of these estimates. The estimated regression equation is

$$Y = A + B (x_i - \bar{x}) , \tag{13}$$

where

$$\bar{x} = \frac{\Sigma x_i}{k} .$$

The method of least squares can be used to estimate regression. This method uses those values of A and b which will minimize the sum

---

4 Brownlee, K. A., Statistical Theory and Methodology in Science and Engineering, p. 274, John Wiley & Sons, Inc. 1960.

of squares of deviations (SSD), between observed values $y_i$ and the predictions $Y_i$ given by putting values of $x_i$ in (13). Therefore, it is desired to minimize

$$SSD = \Sigma (y_i - Y_i)^2 = \Sigma [y_i - A - b (x_i - \bar{x})]^2 . \quad (14)$$

The following relationships result when (14) is differentiated with respect to A and b and the results set equal to zero:

$$\frac{\partial SSD}{\partial A} = -2 \Sigma [y_i - A - b (x_i - \bar{x})] = 0 \quad (15)$$

and

$$\frac{\partial SSD}{\partial b} = -2 \Sigma [y_i - A - b (x_i - \bar{x})](x_i - \bar{x}) = 0. \quad (16)$$

After rearranging terms and using the fact that $\Sigma (x_i - \bar{x}) = 0$, the estimators for $\alpha$ and $\beta$ are

$$A = \frac{\Sigma y_i}{k} = \bar{y} \quad (17)$$

and

$$b = \frac{\Sigma (x_i - \bar{x}) y_i}{\Sigma (x_i - \bar{x})^2} . \quad (18)$$

Since $\Sigma (x_i - \bar{x}) = 0$ and therefore $\bar{y} \; \Sigma (x_i - \bar{x}) = 0$, the numerator of the right hand side of (18) becomes

$$\Sigma (x_i - \bar{x}) y_i - \bar{y} \; \Sigma (x_i - \bar{x}) = \Sigma (x_i - \bar{x}) (y_i - \bar{y}),$$

and an alternative way of expressing b is

$$b = \frac{\Sigma (x_i - \bar{x}) (y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} \quad (19)$$

Since (17) and (18) show that A and b are linear functions of the $y_i$ which are assumed to be independent and have a normal distribution, with variance $\sigma^2$, the expected values and the variances of A and b are

$$E[A] = \Sigma E\left[\frac{y_i}{k}\right] = \frac{1}{k} \Sigma [\alpha + \beta(x_i - \bar{x})] = \alpha, \tag{20}$$

$$E[b] = \frac{\Sigma(x_i - \bar{x}) E[y_i]}{\Sigma(x_i - \bar{x})^2} = \frac{\Sigma(x_i - \bar{x})[\alpha + \beta(x_i - \bar{x})]}{\Sigma(x_i - \bar{x})^2} \tag{21}$$

$$= \frac{\alpha\Sigma(x_i - \bar{x})}{\Sigma(x_i - \bar{x})^2} + \frac{\beta\Sigma(x_i - \bar{x})^2}{\Sigma(x_i - \bar{x})^2} = \beta,$$

$$V[A] = V\left[\frac{\Sigma y_i}{k}\right] = \frac{1}{k^2} \Sigma V[y_i] = \frac{\sigma^2}{k}, \tag{22}$$

and

$$V[b] = V\left[\frac{\Sigma(x_i - \bar{x}) y_i}{\Sigma(x_i - \bar{x})^2}\right] = \frac{\Sigma(x_i - \bar{x})^2}{[\Sigma(x_i - \bar{x})^2]^2} V[y_i] = \frac{\sigma^2}{\Sigma(x_i - \bar{x})^2}. \tag{23}$$

The deviation of an observation $y_i$ from the value predicted by (12) can be written as

$$y_i - \theta_i = (y_i - Y_i) + (Y_i - \theta_i) \tag{24}$$

$$= (y_i - Y_i) + [A + b(x_i - \bar{x}) - \{\alpha + \beta(x_i - \bar{x})\}]$$

$$= (y_i - Y_i) + (A - \alpha) + (b - \beta)(x_i - \bar{x}) .$$

After squaring (24) and summing over i,

$$\Sigma (y_i - \theta_i)^2 = \Sigma (y_i - Y_i)^2 + k (A - \alpha)^2 + (b - \beta) \Sigma (x_i - \bar{x})^2 \tag{25}$$

since the various cross products are zero.

In equation (25), $\Sigma (y_i - \theta_i)^2$ is a sum of squares with k degrees of freedom and is distributed as $\sigma^2 \chi^2 (k)$. It consists of three parts. The parts involving A and b have one degree of freedom each. The third part, $\Sigma (y_i - Y_i)^2$, involves k variables but by setting equations (15) and (16) equal to zero, the degrees of freedom become $(k-2)$.

Since the sum of squares on the left hand side of (25) equals the sum of the three sums of squares on the right hand side of (25) and the degrees of freedom of both sides of (25) are equal, the sums of squares on the right hand side are independent and are distributed as $\sigma^2 \chi^2$ with degrees of freedom $(k-2)$, 1, and 1, respectively. Since $\Sigma (y_i - Y_i)^2$ is distributed as $\sigma^2 \chi^2 (k-2)$ and has mean square $S^2$,

$$E [S^2] = E \left[\frac{\Sigma (y_i - Y_i)^2}{k-2}\right] = \frac{\sigma^2}{k-2} E [\chi^2 (k-2)] = \sigma^2 \tag{26}$$

Therefore $S^2$ is an estimator of $\sigma^2$ and is independent of A and b.

The deviation of $y_i$ from the mean $\bar{y}$ can be written as

$$(y_i - \bar{y}) = (y_i - Y_i) + (Y_i - \bar{y}) \tag{27}$$

After squaring (27) and summing over i,

$$\Sigma (y_i - \bar{y})^2 = \Sigma (y_i - Y_i)^2 + \Sigma (Y_i - \bar{y})^2 \tag{28}$$

where

$$\Sigma (Y_i - \bar{y})^2 = \Sigma [A + b (x_i - \bar{x}) - A]^2 = b^2 \Sigma (x_i - \bar{x})^2 \qquad (29)$$

The sum of the cross products is zero since that sum is in the form of equation (16). The left hand side of (28), $\Sigma (y_i - \bar{y})^2$, is the sum of squares of deviations of observations from the mean, with (k - 1) degrees of freedom. On the right hand side of (28), $\Sigma (y_i - Y_i)^2$ is the sum of squares of deviations of observed values from the estimated line. In (26), it was shown the expected value of the mean square $\Sigma (y_i - Y_i)^2 / (k - 2)$ was $\sigma^2$. The second term $\Sigma (Y_i - \bar{y})^2$ has 1 degree of freedom. Since the sum of squares on the left hand side of (28) equals the sum of two sums of squares on the right hand side of (28) and the degrees of freedom of both sides of (28) are equal, the sums of squares on the right side are independent and are distributed as $\sigma^2 \chi^2$ with degrees of freedom (k - 2) and 1 respectively, as shown in Table IV.

Table IV

Analysis of Variance Table for Deviations from the Mean

| Source of Variance | Degrees of Freedom | Sum of Squares |
|---|---|---|
| Slope of line | 1 | $\Sigma (Y_i - \bar{y})^2$ |
| Deviation of observed value about estimated line | k-2 | $\Sigma (y_i - Y_i)^2$ |
| Total | k-1 | $\Sigma (y_i - \bar{y})^2$ |

If a new observation $y'$ is given and it is desired to find confidence limits for the predicted $x$ corresponding to this $y'$, the estimated regression equation (13) will have to be solved in reverse to obtain

$$b x = b \bar{x} + Y - A \quad . \tag{30}$$

Replacing $x$ by $\hat{x}$ and $y'$ in (30), yields

$$\hat{x} = \bar{x} + \frac{y_i - A}{b} \tag{31}$$

a point estimate of the value $x$ corresponding to the new observation $y'$. The expected value of $y'$ is $\theta$ . Corresponding to this value of $\theta$ is a value of $x$ given by solving the true regression equation (13) for $x$. Calling this value of $x$, $\varepsilon$, then

$$\varepsilon = \bar{x} + \frac{\theta - \alpha}{\beta} \tag{32}$$

or

$$\theta - \alpha - \beta ( \varepsilon - \bar{x}) = 0 \tag{33}$$

Now a new variable $z$ is defined such that

$$z = y' - A - b ( \varepsilon - \bar{x}) \quad . \tag{34}$$

Using the results of (20), (21), and (33), the expected value of $z$ is

$$\begin{aligned} E [z ] &= E [y'] - E [A ] - ( \varepsilon - \bar{x}) E [ b ] \\ &= \theta - \alpha - (\varepsilon - \bar{x}) \beta \\ &= 0 . \end{aligned} \tag{35}$$

Using the results of (22) and (23), the variance of z is

$$V[z] = V[y'] + V[A] + (\varepsilon - \bar{x})^2 V[b] \tag{36}$$

$$= \sigma^2 + \frac{\sigma^2}{k} + (\varepsilon - \bar{x})^2 \frac{\sigma^2}{\Sigma(x_i - \bar{x})^2}$$

$$= \sigma^2 \left[1 + \frac{1}{k} + \frac{(\varepsilon - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}\right] .$$

The random variable z is a linear function of three random normally distributed variables y', A, and b and therefore is normally distributed itself. Thus $\frac{z - 0}{V[z]}$ is Normal (0,1) and, invoking (26) in order to replace $\sigma^2$ by its estimate $S^2$ in (36) and using the relationship in (34),

$$\frac{y' - A - b\,(\varepsilon - \bar{x})}{S\sqrt{1 + \frac{1}{k} + \frac{(\varepsilon - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}}} \sim t\,(k-2) . \tag{37}$$

In order to derive confidence limits for $\varepsilon$, (37) can be inserted for t in the statement $\Pr\{t_1 < t < t_2\} = P_2 - P_1$. Assuming $\varepsilon_2$ is the lower confidence limit, then

$$\frac{y' - A - b\,(\varepsilon_2 - \bar{x})}{S\sqrt{1 + \frac{1}{k} + \frac{(\varepsilon - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}}} = t_2 , \tag{38}$$

where $t_2$ is the $P_2$ point of t with the same degrees of freedom as S, (k - 2).

It is desired to have equation (38) in the form of a quadratic equation in $\varepsilon_2$. Both sides of equation (38) are squared and the results are expanded

$$[\,(y' - A - b\,(\varepsilon_2 - \bar{x})]^2 = t_2^2\, S^2\, (1 + \frac{1}{k} + \frac{(\varepsilon - \bar{x})^2}{\Sigma(x_i - \bar{x})^2}\,) \tag{39}$$

$$[\, b^2 - \frac{t_2^2\, S^2}{\Sigma(x_i - \bar{x})^2}\,]\,\varepsilon_2^2 + 2\,[\frac{t_2^2\, S^2\, \bar{x}}{\Sigma\,(x_i - \bar{x})^2} - b^2\bar{x} - b\,(y' - A)\,]\,\varepsilon_2$$

$$+ (y' - A + b\bar{x})^2 - t_2^2\, S^2\, [1 + \frac{1}{k} + \frac{x^2}{\Sigma(x_i - \bar{x})^2}\,] = 0\,.$$

The solutions for a quadratic equation in the form

$$A\,\varepsilon^2 + B\,\varepsilon + C = 0 \tag{40}$$

are

$$\varepsilon = \frac{-B}{2A} \pm \frac{\sqrt{B^2 - 4AC}}{2A} = \frac{-B}{2A} \pm \sqrt{\frac{(B/2)^2 - AC}{A^2}}\,. \tag{41}$$

In this case

$$-\frac{B}{2A} = \frac{-2\,\{\,[\,t_2^2\, S^2\,\bar{x}/\,\Sigma\,(x_i - \bar{x})^2\,] - b^2\bar{x} - b\,(y' - A)\,\}}{2\,[\,b^2 - t_2^2\, S^2/\,\Sigma\,(x_i - \bar{x})^2\,]} \tag{42}$$

$$= \bar{x} + \frac{b\,(y' - A)}{b^2 - t_2^2\, S^2/\,\Sigma\,(x_i - \bar{x})^2}\,.$$

The numerator of the square of the second term of (41) is

$$(\frac{B^2}{2}) - AC = [\frac{t_2^2\, S^2\, \bar{x}}{\Sigma(x_i - \bar{x})^2} - b^2\bar{x} - b\,(y' - A)\,]^2 \tag{43}$$

$$- [b^2 - \frac{t_2^2 S^2}{\Sigma(x_i - \bar{x})^2} \{ (y' - A + b\,\bar{x})^2 - t_2^2 S^2 [1 + \frac{1}{k} + \frac{\bar{x}^2}{\Sigma(x_i - \bar{x})^2}]\}]$$

$$= [b^2 - \frac{t_2^2 S^2}{\Sigma (x_i - \bar{x})^2} ] (1 + \frac{1}{k}) t_2^2 S^2 + \frac{t_2^2 S^2 (y - A)^2}{\Sigma (x_i - \bar{x})^2} .$$

The second term of (41) is

$$\frac{\sqrt{(\frac{B}{2})^2 - AC}}{A^2} = \frac{t_2 S}{b^2 - \frac{t_2^2 S^2}{\Sigma(x_i - \bar{x})^2}} \sqrt{[b^2 \frac{t_2^2 S^2}{\Sigma (x_i - \bar{x})^2}] (1 + \frac{1}{k} + \frac{(y' - A)^2}{\Sigma (x_i - \bar{x})^2})} \tag{44}$$

When the results of (42) and (44) are substituted in (41) and the negative square root term is used, the solution for the lower confidence limit is

$$\varepsilon_2 = \bar{x} + \frac{b (y' - A)}{b^2 - t_2^2 S^2 / \Sigma (x_i - \bar{x})^2} \tag{45}$$

$$- \frac{t_2 S}{b^2 - t_2^2 S^2 / \Sigma(x_i - \bar{x})^2} \sqrt{[b^2 - \frac{t_2^2 S^2}{\Sigma (x_i - \bar{x})^2} ] (1 + \frac{1}{k} + \frac{(y' - A)^2}{\Sigma (x_i - \bar{x})^2})} .$$

Similarly, the upper confidence limit $\varepsilon$ can be obtained using $t_1$ in place of $t_2$ and its solution is the same as (45) with this change.

In order to simplify (45) somewhat, a new quantity W is defined as

$$W = \frac{t_2^2 S^2}{b^2 \Sigma(x_i - \bar{x})^2} . \tag{46}$$

If W is sufficiently small, ( < 0.1),

$$b^2 - \frac{t_2^2 \, S^2}{\Sigma \, (x_i - \bar{x})^2} - b^2 \, (1 - W) \simeq b^2 \quad . \tag{47}$$

After substituting the approximation from (47) into (45), the approximate confidence limits for $\varepsilon$ are

$$\varepsilon_2 \simeq \bar{x} + \frac{y' - A}{b} - \frac{t_2^2 \, S^2}{|b|} \sqrt{(1 + \frac{1}{k}) + \frac{(y' - A)^2}{b^2 \;\; \Sigma(x_i - \bar{x})^2}} \tag{48}$$

and

$$\varepsilon_1 \simeq \bar{x} + \frac{y' - A}{b} - \frac{t_1 \, S}{|b|} \sqrt{(1 + \frac{1}{k} + \frac{(y' - A)^2}{b^2 \, \Sigma \, (x_i - \bar{x})^2})} \quad . \tag{49}$$

Now, given a new observation y', W is computed and if it is sufficiently small, (48) and (49) will yield the approximate confidence limits for $\varepsilon$.

## LIST OF REFERENCES

Brownlee, K. A., Statistical Theory and Methodology in Science and Engineering, Wiley, 1960.

Stanford Research Institute Technical Report RSSC-TR 4923-5, Communist Terrorist Logistics in Southern Thailand - A Quantitative Analysis, by W. A. Hamberg and C. R. Self, August 1968.

Stanford Research Institute Research Memorandum OAD-RM 4923-5, The Communist Terrorist Organization in Southern Thailand, by D. A. Price, (CONFIDENTIAL), January 1967.

Stanford Research Institute Research Memorandum OAD-RM 4923-8, Notes on the Malayan Strategies and Organization of the Opposing Forces, by P. B. G. Walker, March 1967.

Vogt, Robert L., and Glenn F. Lindsay, A Planning Model for Populace Search and Screening Operations, Naval Postgraduate School, Monterey, California, January 1969.

INITIAL DISTRIBUTION LIST

| | | No. Copies |
|---|---|---|
| 1. | Defense Documentation Center<br>Cameron Station<br>Alexandria, Virginia 22314 | 20 |
| 2. | Library, Code 0212<br>Naval Postgraduate School<br>Monterey, California 93940 | 2 |
| 3. | Chief of Naval Operations (OP96)<br>Department of the Navy<br>Washington, D. C. 20350 | 1 |
| 4. | Professor G. F. Lindsay<br>Department of Operations Analysis<br>Naval Postgraduate School<br>Monterey, California 93940 | 5 |
| 5. | Lieutenant E. F. Sullivan<br>64 Eastland Rd.<br>Jamaica Plain, Massachusetts 02130 | 1 |

Security Classification

# DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

| | |
|---|---|
| 1. ORIGINATING ACTIVITY (Corporate author)<br>Naval Postgraduate School<br>Monterey, California 93940 | 2a. REPORT SECURITY CLASSIFICATION<br><br>2b. GROUP |
| 3. REPORT TITLE<br>Methods for Estimating Present and Future Insurgent Strength | |
| 4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)<br>Master's Thesis; October 1969 | |
| 5. AUTHOR(S) (First name, middle initial, last name)<br>Edward Francis Sullivan | |

| 6. REPORT DATE | 7a. TOTAL NO. OF PAGES | 7b. NO. OF REFS |
|---|---|---|
| October 1969 | 61 | 5 |

| | |
|---|---|
| 8a. CONTRACT OR GRANT NO. | 9a. ORIGINATOR'S REPORT NUMBER(S) |
| b. PROJECT NO. | |
| c. | 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) |
| d. | |

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

| 11. SUPPLEMENTARY NOTES | 12. SPONSORING MILITARY ACTIVITY |
|---|---|
| | Naval Postgraduate School<br>Monterey, California 93940 |

13. ABSTRACT

An estimation of the extent of an insurgency appears to be a prerequisite for determining effective counterinsurgency policies. A method is developed to estimate a confidence interval for present insurgent strength based on consumption of a selected commodity. By examining the anticipated effects of dynamic counter-insurgency programs on additions and deletions to insurgent strength, estimates of future insurgent strength can be attempted. For this purpose, recursive relationships are developed describing changes in insurgent strength which occur with changes in level of a single government activity or multiple governmental activities.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811



Security Classification

A-31408

| 14 KEY WORDS | LINK A | | LINK B | | LINK C | |
|---|---|---|---|---|---|---|
| | ROLE | WT | ROLE | WT | ROLE | WT |
| | | | | | | |

DD FORM 1 NOV 65 1473 (BACK)

S/N 0101-807-6821





A-31409